An effective model for quantification of cell 3Dshape in early embryogenesis

Zelin Li (✉ zelinli6-c@my.cityu.edu.hk)

City University of Hong Kong

Jianfeng Cao

City University of Hong Kong

Zhongying Zhao

Hong Kong Baptist University

Hong Yan

City University of Hong Kong

Research Article

Keywords: Spherical harmonics(SPHARM), Cell shape quantification, Morphological robustness, C. elegans

Posted Date: September 22nd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-911999/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

RESEARCH

An effective model for quantification of cell 3D shape in early embryogenesis

Zelin Li^{1*}, Jianfeng Cao¹, Zhongying Zhao² and Hong Yan¹

*Correspondence:

zelinli6-c@my.cityu.edu.hk

¹Center for Intelligent

Multidimensional Data Analysis
and Department of Electrical
Engineering, City University of
Hong Kong, Kowloon Tong, Hong
Kong

Full list of author information is
available at the end of the article

Abstract

Background: The cell developmental process is featured by fabulous morphogenesis in multicellular organisms. Describing morphological changes quantitatively concretely the way to investigating both intra and inter cell regulations on cell fate. While *Caenorhabditis elegans* has been used as a model for cell development studies for a long time, the research of how cell shape is precisely controlled remains obscure by the lack of methods to model morphological features. Currently, in order to characterize the features of cell shape involved in cell migration and differentiation, there is an increasing demand in analyzing cell shape systematically, especially when many works have contributed to cell reconstruction.

Results: In this work, Spherical harmonics and Principal component analysis integrated Cell Shape quantification Models (SPCSMs) is proposed to represent cell shapes in a low-dimensional shape space. SPCSMs incorporates a complete pipeline to quantify cell shapes and analyze their morphological phenotypes in three dimensional (3D) reconstructions. Based on the framework, we extract biologically meaningful patterns in the lineage of *C. elegans* embryo before 350-cell stage, during which all hypodermis cells deformed like a funnel and can be recognized based on these shape patterns. Finally, SPCSMs is compared with two cell shape representation methods, which substantiates the effectiveness and robustness of our method.

Conclusion: SPCSMs provides a general method to describe shapes in low-dimensional shape space with compact parameters. It can quantify the shapes of cells from single-cell resolution images obtained over one-minute intervals, making it possible to recognize developmental patterns in cell lineages. SPCSMs is expected to be an effective model for biologists to explore the relationships between the shapes of cells and their fates.

Keywords: Spherical harmonics (SPHARM); Cell shape quantification; Morphological robustness; *C. elegans*

Background

Embryogenesis, the development of multicellular organisms from a single-celled zygote [1, 2], involves the division and dynamic morphological and positional changes [3] of cells. Physical interactions between cells, intercellular signaling and forces between cells lead to specific cell fates [4, 5, 6]. Cell shapes are the result of embryo development and growth, which has a key role in cell migration and so are associated with cell differentiation and embryo development [4, 7, 8].

A complete growth map has been determined, presenting the information of the distribution, division and fate of each cell [9]. An automatic system to mark

fluorescently-labeled *C. elegans* embryos has been developed [10] and this facilitates segmentation with deep learning [11] with a labeled nucleus which have higher segmentation result.

Although imaging technology can provide clear pictures, the dynamic process of changes in cell shapes has not been clearly defined in early embryonic development. As cell fate is mostly determined by differential gene expression [1, 2], and partly induced through signaling interactions via cell-cell contacts [8, 12], cell shapes could be used to measure or predict the surface contacts, interactive forces and signaling between cells and their morphological types, functional features, or relative positions.

We propose to analyze cell shapes using a primary model, "Spherical harmonics and Principal component analysis integrated Cell Shape quantification Models" (SPCSMs), which was developed to examine the embryogenesis of *C. elegans* up to the 350-cell stage but it would be effective on other multicellular organisms. *C. elegans* are transparent and their development has been studied in *vivo* at single-cell resolution, allowing examination of new or existing cell-cell interactions and visualization of cell division in real-time. It has long been used as a model organism, and its genetics and genomics, signal transduction, and evolution are well studied [13, 14].

Three dimensional reconstructions of cell shapes can be generated with deep learning models [15] or other image processing methods [7, 16]. To analyze cell shapes by 3D object representation methods, a quantification of the shape is needed, such as outline extraction and outline-based PCA (outline PCA) [17], spherical harmonic (SPHARM) modeling[18] or surface curvature descriptors[19]. This quantification of a shape should contain as much morphological information as possible.

We aim to identify the regulation of clusters of cells by their shapes during embryo embryogenesis. The mechanical forces between interacting cells, which catalyze biological patterns[3, 8] and physical timing constraint conditions[4] in early embryogenesis, could be derived from the model quantifying the shapes of cells.

Results

SPCSMs integrates the orthogonal completeness and rotation invariance of SPHARM and PCA to provide a improved parameterization. The quantitative evaluation shows that SPCSMs performs better at quantifying and representing cell shape than SPHARM in low-dimensional space. It shows rotational invariance which outline-based methods do not have. Analysis of cell shape lineage that conducted with SPCSMs revealed clustering patterns in embryogenesis. The degree used in this section is clockwise and right-hand cartesian coordinate is used when rotation experiment is applied.

Performance at Representing Segmented Cells

We compared the effectiveness Table. 1 which represents the 3D surface reconstruction by outlining the PCA, SPHARM, and SPCSMs. The outline process used the sampling points on the object's surface to visualize the reconstructed object in 3D and quantitatively analyze the errors (Figs 1, 2 and 3). The comparison experiment is conducted to verify the representation effectiveness of these three methods. The

numbers of sample points of outline extraction (direct sampling points) range from 50 to 648. The degree (numbers of principal components) of outline PCA (by applying PCA after directly sampling points) range from 5 to 24, and the degrees of SPHARM and SPCSMs both range from 6 to 25 (the number of coefficients is the square of degree number). The median of all cells with the same degree representation is considered to be the error of this degree. The error curves are power-law. From the error curves of SPHARM, degree 25 was chosen as the number of original sample points for SPHARM. The representation error results showed that the accuracy of SPCSMs can become higher than SPHARM as the embryo develops. The average error became less than 0.1 from degree 17 (289 coefficients) for SPCSMs and from degree 21 (441 coefficients) for SPHARM. Outline PCA was the worst.

Rotation Invariance of SPCSMs

SPCSMs integrates the SPHARM and PCA. It is invariant on rotation so that the sum of the coefficients of a 3D object representation with the same degree do not change when rotated. Experiments were conducted on regular polyhedrons and cells with SHTools[20]. The distance between vectorization $\vec{x}_i$ of two cells from SPCSMs can be used to learn the similarity of two cells in embryonic development.

Regular Polyhedrons For each cube and regular tetra-, octa-, dodeca-, and icosahedra, five random radius pairs in latitude and longitude were generated and the polyhedra were radially rotated. Their rotated shape, spectrum curves, and 2D spectra were plotted. Through this experiment, the rotation invariance can be proven in spectrum curve of SPHARM. The sum of coefficients with the same degree does not change when 3D objects are rotated Fig.6. Six 3D plots of original and rotated cubes are shown in Fig.4. The SPHARM coefficients in a 2D spectra are given in Fig.5. Cell registration can be evaluated in SPCSMs by comparing the similarity of the spectrum curves of two cells using the euclidean distance between two SPCSMs quantification vectors, $\vec{c}$, of cells.

Biological Dataset The rotation invariance of SPCSMs has been proven by regular polyhedrons experiments. To further illustrate the rotational invariance of SPCSMs, a Caaaa cell from six sequential time-points (108-113) at the 350-cell stage in the mitosis process were used to obtain their SPHARM spectrum curve and 2D spectra from SPHARM. As shown (Figs. 7, 8, and 9), the 2D spectra (the sum of the same degree coefficients of SPHARM) change smoothly and continuously as the cell shapes change. Because these cell shapes are almost the same and their 2D spectra curve are highly similar. This property can be automated for registration and matching of cell shapes to place cells in comparable vector space.

Component Shape of SPCSMs

SPCSMs is an effective model in cell lineage analysis. Shape deformation can be seen and cell development regulations can be derived from the results of SPCSMs (cell shape lineage analysis in Fig. 10). SPCSMs result are calculated from the eigenvectors of the covariance matrix of the dataset (about 250,000 cells and 676 coefficients for each cell). The top 12 principle components from the covariance matrix provide 82.5% of the differentiation in the whole dataset during the 350-cell stage of

embryonic development. The projection of SPHARM $\vec{b}$, the vectorization result of SPCSMs $\vec{c}$, in the direction of the first component contains 17.2% of contribution to the variation in the dataset.

Fig. 10 shows the top principle component and the top 6 components shape description which are given in Table. 2. All 12 component shapes can be found in Additional files. Every component reflects part of the original cell shape. We can also find out that there are some regulations of the first component of cell shape lineage analysis in embryonic development. While the first component is the most significant observation object with the least noise in Table. 2, we know that the sunken is produced by the depressing force from surrounding interactive cells by observing its shape in Fig. 10. From the shape of the first component, the negative value of it means the vertical force is exerted inward by surrounding cells to this cell. Otherwise, the force is horizontal. A few physical information is derived by shape quantification results which can help analyze cell shape lineage and predict cells movement in our future work.

Cell Lineage Analysis

Hypodermis Clustering

The major discovery is that the cell shape analysis can cluster the hypodermis (hypodermal cells, that would become skin tissue of *C. elegans*) cells. The main cell shape lineage tree based on the first component values (the first term in $\vec{c}$, it contains 17.2% shape variation in dataset) of the SPCSMs vectorizations, $\vec{c}$, of all cells in embryonic development (each cell is represented by SPCSMs giving its $\vec{c}$ and its first element is this vector $\vec{c}$ of this cell in lineage tree) is shown in Fig. 11. The first element value of $\vec{c}$ representing 17 embryos showed this clustering phenomenon. The hypodermis cells, the posterior two granddaughters of ABa, anterior two granddaughters of ABp and daughters of C, are clustered in cell shape lineage analysis Fig. 11. The hyp 7 generated form P lineage (Caaaa, Caaap, Cpaaa, Cpaap, Cpapa) have large negative values and cluster in an obvious way. Vertical pressure and forces are likely to generate their shapes. They are ectoderm and distribute at the outer layer of the embryo differentiating into skin tissues of *C. elegans*. For other hypodermis, the sub lineage daughter cells (ABarpaap, ABarpapapa, ABarpapp lineage) of ABa and (ABplaaaap, ABplaaapa, ABplaapp lineage) of ABp cells also cluster in the lineage tree and these features would continue from 170-cell stage embryo. 80% the skin cells cluster in the lineage tree because of they are pressing hard inside vertically or horizontally. Now this clustering phenomenon of hypodermal cells is recognized by SPCSMs. These cells receive some signaling from their ancestors and surrounding cells because of their particular shapes. More information and biological clues can be discovered with SPCSMs in the future. These patterns have been observed from all 17 embryos in our dataset. The result figures of their cell shape lineage tree can be found in Additional files. It's reproducible in all 17 embryos.

Hypodermis Shape Display

From a physical standpoint, inside vertical and horizontal forces (funnel-like shape in Fig. 10 of cells are shown in Fig. 11) are presented during the determination of the fate of cells. The red and blue are clustering, which means a type of cells are usually

depressed by an instant force (constant values in Fig. 11 during hypodermis cell life cycle) with unchanged direction until cells complete mitosis division and functional differentiation. This force would possibly influence their deformation(shape) and movement in 350-cell stage embryonic embryogenesis. That is why we can find this clustering phenomenon with SPCSMs. The results of SPCSMs contain not only the spatial shape information but also the physical information of the cells.

Deformation Analysis

SPCSMs result $\vec{c}$, vectorization of cell, transform cell shape into the shape space of our dataset. The euclidean (L-2 norm) distance of two $\vec{c}$ means the difference degree of two cell shapes. A deformation degree lineage tree is provided in Fig. 12, the cell would become sphere shape before it separate. We can also observe that hypodermis cells deformed severely. The lineage cell analysis verifies the effectiveness of SPCSMs. The result are consistent with biological law.

Discussion

SPCSMs vectorize cell shapes with less noise than original segmented cell data and our proposed methodology to analyze the shapes of cells from these vectors can be applied to discover cell types and represent cell shapes. By integrating SPHARM and PCA, SPCSMs had the property of rotation invariance. Existing shape quantification methods were not designed for biological research. SPCSMs provides a good vectorization method to quantify cell shapes. Using SPCSMs helps biologists find meaningful biological patterns (recognize hypodermis cells and deformation degree of cells) in embryonic development Fig. 11. The quantification and calculation procedures can be used for cell shape lineage analysis. Before the 350-cell stage of embryonic embryogenesis, the relationship between the weight of each component of SPCSMs and cell fate was calculated and plotted in lineage trees. The fates of the hypodermis cells are highly associated with deformation in their shapes.

Outline PCA, SPHARM and SPCSMs

Outline PCA is a common 3D object representation method that extracts sampling points on a membrane (surface) and can represent cell shapes by this group of points. Theoretically, most of the features and curvature could be represented by these points. However, although outline PCA reduces the dimension from outline extraction, more than ten thousand points are likely to be needed for one cell, making it intractable for most embryogenesis datasets. The performance curves of three methods are plotted in Fig. 1 (outline PCA), Fig. 2 (SPHARM) and in Fig. 3 (SPCSMs). SPCSMs shows the best performance.

Cell Type and Cell Fate

Using segmented 3D images of cells, K-means clustering was applied to cluster the cells based on the SPCSMs results analysis. Regular shapes were clustered with the SPCSMs results and the 3D object could be grouped to correct clusters (Additional Files with K-means clustering algorithm[21]). This ideal situation (shapes rotate in regular radius) shows the hyper-parameter K can cluster these objects and could be applied to cell shape quantification results. Theoretically, we could group same type

of cells into one cluster with the same shape of the results from SPCSMs. However, we do not know how many clusters, and so the perfect hyper-parameters K in the K-means algorithm, are in the high-dimensional vector space. We can not know the perfect hyper-parameters K in the K-means algorithm. Types of cells could be clustered into several groups if we integrate results of SPCSMs, curvatures of cell shapes and surface force analysis. This might help improve cell mechanical models in future work.

By cell lineage analysis, in *C.elegans*, the hypodermis (future skin tissues) cell type (some vent gang and ring gang cells would recognized as hypodermis cells) can be recognized by SPCSMs but neuronal, muscle, and intestinal cells were not. SPCSMs focuses on the global dataset and neglects local changes on the cell membrane surface. Curvature and the influence of physical forces should be considered in future work when we continue to working with this model.

Conclusion

This work presents an effective shape model focusing on representation and quantification of cells. SPCSMs performs better than outline PCA, and SPHARM with the same number of coefficients. SPCSMs vectorizes cell shapes and reduces noise with PCA. We propose a methodology to analyze the results of SPCSMs and use it to find biological patterns in 350-cell stage *C. elegans* embryos. Finding patterns in the shapes of cells in a developing embryo is technically challenging. Still our methodology is effective for large-scale analysis of cell shapes and exploration of hypodermis cells in lineage tree. Future work would include verifying and exploring the patterns in biological shapes, analyzing cell movement, and building mechanical model on *C. elegans*.

Methods

SPCSMs provides a feasible pipeline to quantify cell shape and cluster cells. All cells are translated to the same coordinate system with membrane surface data generated by CShaper[11] so that sample points can collected on them. The spherical harmonics (SPHARM) coefficients are calculated (Eq.4) and the membrane surface is transformed into SPHARM space (Eq.3), as the frequency domain of a Fourier transform. Several principal components, whose directions have the biggest variance, are selected and the PCA transform of matrices of cells' SPHARM coefficients can be derived. Cell shapes are clustered using the quantification result of SPCSMs.

Data

The multicellular organism, *C. elegans* was used. 2D image stacks were processed to 3D objects with CShaper[11, 22]. In embryonic development, the number of cells increases as cells divide over embryo development. The original three dimensional cell surface points set are produced by CShaper[11]. That is the input of SPCSMs.

Map 3D Surface

Spherical harmonics transform (SPHARM) needs to sample the 3D surface and the number of sample points needs to meet the sampling theorem conditions[23]. We get a compromised hyperparameter condition from the error curve in Fig. 3 from

the error curve of the result of SPCSMs. 17 degrees were used in SPCSMs in the cell lineage analysis and clustering experiment. There are $289 = 17^2$ dimensions after the SPHARM transform which requires at least $34 = 17 \times 2$ sample intervals in latitude and longitude. Therefore, we need to sample $2312 = 34^2 \times 2$ points on a 3D object's surface, which are then mapped to a 2D grid. A set of (x, y, z) points in cartesian coordinates would be transformed to a group of (r, θ, ϕ) points in spherical coordinates; θ, ϕ are the latitude and longitude in 3D and r is the distance from the centroid. On the 2D grid, longitude ϕ is the abscissa and latitude θ is the ordinate while r is the value for each point on the 2D grid (Fig. 13).

Cell Shape Quantification

To measure the representation error of outline PCA and SPCSMs, distances from the cells' centroid of several random sampling points on the reconstructed surfaces of outline PCA and SPCSMs were calculated. The criterion is to compare their error with the same dimension after vectorization in $\vec{x}_i$ (with the same number of parameters). $\vec{x}_i$ means the vectorization (quantification) result of i-th cell in the dataset. For example, outline PCA samples 450 points on the original surface and then keeps the first 225 principal components only. The dimension of the outline PCA $\vec{a}$ would be 225. SPHARM would take degree 15 and there would be 225 coefficients in the SPHARM vectorization, $\vec{b}$. 225 coefficients are in the SPCSMs representation of vector $\vec{c}$. Then the results of these three methods are comparable. $\vec{a}, \vec{b}, \vec{c}$ are used to represent the vectorization results of outline PCA, SPHARM and SPCSMs.

Outline PCA

The process of outline PCA has the same calculation steps as SPCSMs. The dataset matrix is constructed by vectorizing every cell (Eq.7). The vector $\vec{x}_i$ is the distances r from the centroid of a series of sample points on the original surface, rather than a series of parameters $\vec{f}_i$ from an SPHARM transform. The reconstruction process of outline PCA in 3D space is finished with the same steps as the PCA process (Eq.11) so that the result would be a series of points on the surface with a regular interval in latitude and longitude in spherical coordinates. Spherical coordinates are commonly used as cell shapes are approximately spherical in the embryogenesis of *C. elegans*.

Spherical Harmonics

Spherical harmonics (SPHARM) is a form of Fourier series, known as a Fourier spherical harmonic function or Laplace's spherical harmonics [24, 18]. Its transform can be applied to parameterize a 3D object, especially a spherical object. Unlike Fourier series, another set of orthogonal functions, $Y_l^m(\theta, \phi)$, is used rather than a set of *sin* and *cos* functions. They can be viewed as the angular portion of solutions to Laplace's equation in 3D [25]. The SPHARM orthogonal basic function

$$Y_l^m(\theta, \phi) = \sqrt{\frac{2l+1}{4\pi} \frac{(l-m)!}{(l+m)!}} P_l^m(\cos\theta) e^{im\phi} \quad (1)$$

where $P_l^m(\cos\theta)$ are the associated Legendre polynomials defined by the differential equation

$$P_l^m(\omega) = \frac{(-1)^m}{2^l l!} (1 - \omega^2)^{\frac{m}{2}} \frac{d^{l+m}}{d\omega^{l+m}} (\omega^2 - 1)^l \quad (2)$$

Equation (2) enables the SPHARM transform, using $Y_l^m(\theta, \phi)$ to represent the surface(membrane) equation or distribution. After mapping all the sample points $\mathbf{r}(\theta, \phi) = (x(\theta, \phi), y(\theta, \phi), z(\theta, \phi))$ on the surface, we assume that the surface equation $f(\theta, \phi)$ represents the cell surface curve in a spherical coordinate system [26]. The definition of the Laplace spherical harmonics $Y_l^m : \mathbf{S}^2 \rightarrow \mathbb{C}$, Y_l^m forms a complete set of orthonormal basic functions in Hilbert space [20]. Thus, on the surface of a unit sphere $\mathbf{S}^2$, the surface equation $f : \mathbf{S}^2 \rightarrow \mathbb{C}$ is expanded and represented as

$$f(\theta, \phi) = \sum_{l=0}^L \sum_{m=-l}^l f_l^m Y_l^m(\theta, \phi) \quad (3)$$

where the f_l^m refer to

$$f_l^m = \int_{\Omega} f_l^m Y_l^m(\theta, \phi) d\Omega \quad (4)$$

$$= \int_0^{2\pi} d\phi \int_0^\pi d\theta \sin\theta f(\theta, \phi) Y_l^m(\theta, \phi) \quad (5)$$

Equation (4) is the SPHARM expansion calculation method. f_l^m are calculated from a Fourier integral function rather than a standard least-squares estimation [18, 25]. The integral equation (4) gives more accurate SPHARM coefficients from the sampling points on the membrane surface. Consider L as the SPHARM degree, a 3D cell object would be represented by a vector

$$\vec{f}_i = [f_0^0, f_1^{-1}, f_1^0, f_1^1, f_2^{-2}, f_2^{-1}, f_2^0, f_2^1, f_2^2, \dots, f_L^{-L}, f_L^{-(L-1)}, \dots, f_L^{(L-1)}, f_L^L]^T \quad (6)$$

where $\vec{f}_i$ represents the i -th cell object in the data, L means the max degree used in the SPHARM transform.

If we have N cells objects in our data, a $(L+1)^2 \times N$ matrix $\mathbf{F}$ is generated as

$$\mathbf{F} = [\vec{f}_1, \vec{f}_1, \vec{f}_2, \dots, \vec{f}_{N-1}, \vec{f}_N] \quad (7)$$

representing the whole dataset. The SPHARM transform is complete for the whole dataset and gives the linear shape representation for all 3D cell objects.

Principal Component Analysis

Using the SPHARM expansion Equation (3), the surface curve is vectorized to a set of SPHARM coefficients, a vector in SPHARM space. Although a representation is found for the membrane surface, the size of the dataset is too large for the following analysis. An algorithm needs to be applied to reduce the vector dimension, and PCA is commonly used.

PCA is a widely used statistical learning method. It can be considered as a kind of transformation. The original space is transformed into an orthogonal space system. The representation of the object does not change, but the direction and position of the space system does. PCA is used to find the direction and position to rotate and translate the space system.

The principal components would be sorted by their variance. The first principle component lies in the scalar projection of the data with the greatest variance. The vector $\vec{z}_i$ of SPCSMs (Eq.9) represented in PCA space would have the same dimension as the original space. There are two ways to calculate the explained variance and principal components: using Singular Value Decomposition (SVD) or calculating eigenvectors from the variance matrix [27]. Using SVD is more efficient when the sample N is large (Eq.7). SVD was implemented with scikit-learn[28]. Consider there are N cells in the data, from Eq.7, we can derive principal matrices

$$\mathbf{P} = [\vec{p}_1, \vec{p}_2, \dots, \vec{p}_k] \quad (8)$$

where k is the dimension of PCA, the dimension to reduce to, and $\vec{p}_i \in \mathbb{R}^d$, while $d = (L + 1)^2$.

For a transformed shape, an SPHARM coefficient vector $\vec{f}_i$ as in Eq.6 (for convenience $\vec{x}_i$ is used rather than $\vec{f}_i$ in the following discussion), its low dimensional transformation based on k principal components (Eq.8) would be

$$\vec{z}_i = (\mathbf{P})^{-1} \cdot (\vec{x}_i - \vec{\mu}) \quad (9)$$

where $\vec{z}_i \in \mathbb{R}^k$ and $\vec{x}_i, \vec{\mu} \in \mathbb{R}^d$, while $\vec{\mu}$ is the mean vector (PCA mean value) in the whole SPHARM dataset (Eq.7).

To calculate the inverse of matrix $\mathbf{P}$ when it is not a square matrix, QR decomposition [28, 29] is applied to calculate the $(\mathbf{P})^{-1}$. $\mathbf{P}$ is decomposed into an orthogonal matrix $\mathbf{Q}$ and an upper triangular matrix $\mathbf{R}$, and minimizing the norm $\|\mathbf{P} \cdot \vec{z}_i - \vec{x}_i\|$ solves the problem $\vec{x}_i = \mathbf{P} \cdot \vec{z}_i$. Therefore, the Eq.9 would be modified as

$$\hat{\vec{z}}_i = (\mathbf{R})^{-1} \cdot (\mathbf{Q}^T \cdot (\vec{x}_i - \vec{\mu})) \quad (10)$$

PCA is an estimation method, which means there would be more errors when using smaller k . The backward calculation for $\hat{\vec{x}}_i$ would be

$$\hat{\vec{x}}_i = \mathbf{P} \cdot \hat{\vec{z}}_i \quad (11)$$

The SPCSMs representation vector for every cell in the whole dataset with N sample cells in a time sequence is obtained. SPCSMs is a method for vectorizing cell shapes and reducing noise in the whole dataset. Integrated PCA does not eliminate noise, but can reduce noise by projecting the representation vector $\vec{x}_i$ in the major direction. The first projection direction has the largest explained variance, which means it contains most of the information in the dataset. The noise is usually white noise and would be more prominent in the latter components, rather than the top components.

Appendix

Funding

This work is supported by Hong Kong Innovation and Technology Commission (Project CIMDA) and Hong Kong Research Grants Council (Project CityU 11204821). The funding body did not play any role in the design of the study, the collection, analysis, and interpretation of the data and in the writing of the manuscript.

Abbreviations

C. elegans: *Caenorhabditis elegans*; SPHARM: Spherical harmonics; SPCSMs: SPHARM and PCA integrated Cell Shape quantification Model;

Availability of data and materials

The dataset and more experiments supporting the conclusions of this article are available at the additional files[30].

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

ZZ and HY initiated the project. ZL and HY designed the methodology. JC provided segmented data and gave important advice. ZZ provided the original dataset and biological advice. All authors revised the manuscript. All authors read and approved the final version of the manuscript.

Author details

¹Center for Intelligent Multidimensional Data Analysis and Department of Electrical Engineering, City University of Hong Kong, Kowloon Tong, Hong Kong. ²Department of Biology, Hong Kong Baptist University, Kowloon Tong, Hong Kong.

References

- Hunt-Newbury, R., Viveiros, R., Johnsen, R., Mah, A., Anastas, D., Fang, L., Halfnight, E., Lee, D., Lin, J., Lorch, A., et al.: High-throughput in vivo analysis of gene expression in *Caenorhabditis Elegans*. *PLoS biology* **5**(9), 237 (2007)
- McKay, S., Johnsen, R., Khattri, J., Asano, J., Baillie, D., Chan, S., Dube, N., Fang, L., Goszczynski, B., Ha, E., et al.: Gene expression profiling of cells, tissues, and developmental stages of the nematode *C. elegans*. In: Cold Spring Harbor Symposia on Quantitative Biology, vol. 68, pp. 159–170 (2003). Cold Spring Harbor Laboratory Press
- Li, X., Zhao, Z., Xu, W., Fan, R., Xiao, L., Ma, X., Du, Z.: Systems properties and spatiotemporal regulation of cell position variability during embryogenesis. *Cell reports* **26**(2), 313–321 (2019)
- Fickentscher, R., Struntz, P., Weiss, M.: Setting the clock for fail-safe early embryogenesis. *Physical Review Letters* **117**(18), 188101 (2016)
- Lieber, A.D., Yehudai-Resheff, S., Barnhart, E.L., Theriot, J.A., Keren, K.: Membrane tension in rapidly moving cells is determined by cytoskeletal forces. *Current biology* **23**(15), 1409–1417 (2013)
- Fickentscher, R., Struntz, P., Weiss, M.: Mechanical cues in the early embryogenesis of *Caenorhabditis Elegans*. *Biophysical journal* **105**(8), 1805–1811 (2013)
- Du, C.-J., Hawkins, P.T., Stephens, L.R., Bretschneider, T.: 3D time series analysis of cell shape using laplacian approaches. *BMC bioinformatics* **14**(1), 1–16 (2013)
- Guignard, L., Fiúza, U.-M., Leggio, B., Laussu, J., Faure, E., Michelin, G., Biasuz, K., Hufnagel, L., Malandain, G., Godin, C., et al.: Contact area-dependent cell communication and the morphological invariance of ascidian embryogenesis. *Science* **369**(6500) (2020)
- Rose, L.S., Kemphues, K.J.: Early patterning of the *C. elegans* embryo. *Annual review of genetics* **32**(1), 521–545 (1998)
- Murray, J.I., Bao, Z., Boyle, T.J., Waterston, R.H.: The lineaging of fluorescently-labeled *Caenorhabditis Elegans* embryos with starrynite and acetree. *Nature protocols* **1**(3), 1468–1476 (2006)
- Cao, J., Guan, G., Ho, V.W.S., Wong, M.-K., Chan, L.-Y., Tang, C., Zhao, Z., Yan, H.: Establishment of a morphological atlas of the *Caenorhabditis Elegans* embryo using deep-learning-based 4D segmentation. *Nature communications* **11**(1), 1–14 (2020)
- Krishnaswamy, S., Spitzer, M.H., Mingueneau, M., Bendall, S.C., Litvin, O., Stone, E., Pe'er, D., Nolan, G.P.: Conditional density-based analysis of t cell signaling in single-cell data. *Science* **346**(6213) (2014)
- Bucher, E.A., Seydoux, G.: Gastrulation in the nematode *Caenorhabditis elegans*. In: *Seminars in Developmental Biology*, vol. 5, pp. 121–130 (1994). Elsevier
- Epstein, H.F., Shakes, D.C.: *Caenorhabditis Elegans: Modern Biological Analysis of an Organism*. Academic Press, ??? (1995)
- He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778 (2016)
- Phan, S., Boassa, D., Nguyen, P., Wan, X., Lanman, J., Lawrence, A., Ellisman, M.H.: 3D reconstruction of biological structures: automated procedures for alignment and reconstruction of multiple tilt series in electron tomography. *Advanced structural and chemical imaging* **2**(1), 1–18 (2016)

17. Pincus, Z., Theriot, J.: Comparison of quantitative methods for cell-shape analysis. *Journal of microscopy* **227**(2), 140–156 (2007)
18. Shen, L., Farid, H., McPeck, M.A.: Modeling three-dimensional morphological structures using spherical harmonics. *Evolution: international journal of organic evolution* **63**(4), 1003–1016 (2009)
19. Czerniawski, T., Nahangi, M., Haas, C., Walbridge, S.: Pipe spool recognition in cluttered point clouds using a curvature-based shape descriptor. *Automation in Construction* **71**, 346–358 (2016)
20. Wieczorek, M.A., Meschede, M.: SHTOOLS: Tools for working with spherical harmonics. *Geochemistry, Geophysics, Geosystems* **19**(8), 2574–2592 (2018)
21. Krishna, K., Murty, M.N.: Genetic k-means algorithm. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* **29**(3), 433–439 (1999)
22. Cao, J., Wong, M.-K., Zhao, Z., Yan, H.: 3DMMS: robust 3D membrane morphological segmentation of *C. elegans* embryo. *BMC bioinformatics* **20**(1), 1–13 (2019)
23. Driscoll, J.R., Healy, D.M.: Computing fourier transforms and convolutions on the 2-sphere. *Advances in applied mathematics* **15**(2), 202–250 (1994)
24. Brechbühler, C., Gerig, G., Kübler, O.: Parametrization of closed surfaces for 3D shape description. *Computer vision and image understanding* **61**(2), 154–170 (1995)
25. Ruan, X., Murphy, R.F.: Evaluation of methods for generative modeling of cell and nuclear shape. *Bioinformatics* **35**(14), 2475–2485 (2019)
26. Laga, H., Takahashi, H., Nakajima, M.: Spherical wavelet descriptors for content-based 3D model retrieval. In: *IEEE International Conference on Shape Modeling and Applications 2006 (SMI'06)*, pp. 15–15 (2006). IEEE
27. Wold, S., Esbensen, K., Geladi, P.: Principal component analysis. *Chemometrics and intelligent laboratory systems* **2**(1–3), 37–52 (1987)
28. Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Mueller, A., Grisel, O., Niculae, V., Prettenhofer, P., Gramfort, A., Grobler, J., Layton, R., VanderPlas, J., Joly, A., Holt, B., Varoquaux, G.: API design for machine learning software: experiences from the scikit-learn project. In: *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*, pp. 108–122 (2013)
29. Goodall, C.R.: *13 COMPUTATION using the QR decomposition* (1993)
30. Additional files and figures (2021). <https://1drv.ms/u/s!Ath3f5ykGji3njfgvv9qdE6R4j1s?e=5WJRL8> Accessed 10 Aug 2021

Figures

Figure 1 Outline PCA error curve The error boxplot of outline PCA. Each point in this figure represents the average error (the y axis) of a single cell, which is generated from the errors of 1000 random points on the cell surface. The x axis represents how many coefficient are used. The number of coefficients is equal to the square of number of degree.

Figure 2 SPHARM error curve The error boxplot of SPHARM.

Figure 3 SPCSMs error curve The error boxplot of SPCSMs.

Figure 4 Cube Rotation 3D plots of a rotated cube (a regular polyhedron). The rotation degree is $\pi/4$ in either latitude or longitude. The cube rotates $\pi/8$ radians clockwise along the x axis and $\pi/8$ along z axis in each sub figure. The degree in figure is clockwise and right-hand cartesian coordinate is used.

Figure 5 Cube Rotation 2D Spectra The 2D spectra of the six rotated cubes in Fig. 4.

Figure 6 Cube Rotation Spectrum Curve The representation results of SPHARM corresponding to six cubes in Fig. 4. The curves are totally same.

Figure 7 Time-lapse cell shape changes of Caaaa.

Figure 8 Shape representations of cell Caaaa based on SPHARM

Figure 9 Caaaa tp:110 Degree Curve The 2D spectra of the six Caaaa cells. They have similar spectrum curves.

Figure 10 The first principal component shape with different weights of SPCSMs First component explains 17.4% of the differentiation. SPCSMs represents cell shape with the top 24 principal components. This figure shows the first principal component shape with different weight values (24 components constitute each cell with 24 different coefficients for each shape).

Figure 11 Overview of the principal component in *C. elegans* lineage First component explains 17.4% of the differentiation. The reproducible pattern in the embryonic development is exemplified by Embryo03 in [11].

Figure 12 Statistical analysis on cell deformation This figure is analyzed based on 250,000 cells in the dataset [11]. While the average shape of these cells is a sphere, the difference between each cell and the average shape elucidates the deformation degree of this cell. Here colors green and red indicate regular and irregular cells, respectively.

Figure 13 An example of representing 3D shape in 2D matrix

Table 1 Error Median Table

The error median of degree of outline PCA, SPHARM and SPCSMs.

Degree	outline PCA	SPHARM	SPCSMs
8	0.397210	0.177560	0.167752
11	0.299744	0.138649	0.118901
15	0.224611	0.114582	0.100925
16	0.207957	0.110431	0.098748
19	0.181356	0.102087	0.095154
20	0.181477	0.099930	0.094517
24	0.153549	0.092803	0.092324

Table 2 Analysis of SPCSMs Result Shape

The top 6 components of SPCSMs are placed in additional files. The sum of their variance is 67.4%				
Index	Variance	Pattern Extent	Minus Deformation	Plus Deformation
1	17.2%	always	vertical depress inside	horizontal depress inside
2	14.3%	sometimes	y-axis sunken	x-axis depress inside
3	11.9%	seldom	elongated and clockwise	counterclockwise and stretch
4	11.2%	seldom	severely twist and counterclockwise	severely twist and clockwise
5	9.76%	sometimes	capsule shape and counterclockwise	capsule shape and clockwise
6	2.93%	sometimes	protrusion towards right	protrusion towards left

Tables

Additional Files

Additional file 1 — Regular polyhedrons SPCSMs result

This is a .csv file that contains the SPCSMs results of 20,000 regular polyhedra. The row index has three parts: polyhedron type, latitude rotated rad, longitude rotated rad. The polyhedron type map is 0:'unit-regular-tetrahedron', 1:'unit-cube', 2:'unit-regular-octahedron', 3:'unit-regular-dodecahedron', 4:'unit-regular-icosahedron'.

Additional file 2 — Cluster Result of SPCSMs

This is the cluster result by applying the K-means algorithm on SPCSMs results of 20,000 regular polyhedra. The same cluster number means they're clustered to one group but the number order has no meaning. The same polyhedra are clustered in one correct group.

Additional file 3 — Biological data SPCSMs components

The SPCSMs components from the whole dataset (8 embryos). These components were used to calculate the $\vec{z}_i$.

Additional file 4 — SPCSMs result of one embryo

The vectorization result from SPCSMs of one embryo. This is the significant data used while analyzing the cell shape lineage and time-lapse morphological development.

Additional file 5 — SPCSMs shape display

The result of SPCSMs are displayed in this .zip file. Each component shape is shown in one figure. The order indicates how much information they contain. The negative number or positive number above sub-figure means what is the component shape would look like in SPCSMs result.

Figures

Figure 1

Outline PCA error curve The error boxplot of outline PCA. Each point in this figure represents the average error (the y axis) of a single cell, which is generated from the errors of 1000 random points on the cell surface. The x axis represents how many coefficient are used. The number of coefficients is equal to the square of number of degree.

Figure 2

SPHARM error curve The error boxplot of SPHARM.

Figure 3

SPCSMs error curve The error boxplot of SPCSMs.

Figure 4

Cube Rotation 3D plots of a rotated cube (a regular polyhedron). The rotation degree is $\pi/4$ in either latitude or longitude. The cube rotates $\pi/8$ radians clockwise along the x axis and $\pi/8$ along z axis in each sub figure. The degree in figure is clockwise and right-hand cartesian coordinate is used.

Figure 5

Cube Rotation 2D Spectra The 2D spectra of the six rotated cubes in Fig. 4.

Figure 6

Cube Rotation Spectrum Curve The representation results of SPHARM corresponding to six cubes in Fig. 4. The curves are totally same.

Figure 7

Time-lapse cell shape changes of Caaaa.

Figure 8

Shape representations of cell Caaaa based on SPHARM

Figure 9

Caaaa tp:110 Degree Curve The 2D spectra of the six Caaaa cells. They have similar spectrum curves.

Figure 10

The first principal component shape with different weights of SPCSMs First component explains 17.4% of the differentiation. SPCSMs represents cell shape with the top 24 principal components. This figure shows the first principal component shape with different weight values (24 components constitute each cell with 24 different coefficients for each shape).

Figure 11

Overview of the principal component in *C. elegans* lineage First component explains 17.4% of the differentiation. The reproducible pattern in the embryonic development is exemplified by Embryo03 in [11].

Figure 12

Statistical analysis on cell deformation This figure is analyzed based on 250,000 cells in the dataset [11]. While the average shape of these cells is a sphere, the difference between each cell and the average shape elucidates the deformation degree of this cell. Here colors green and red indicate regular and irregular cells, respectively.

Figure 13

An example of representing 3D shape in 2D matrix

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [AdditionalFiles.zip](#)